Deep Intelligent Network for Device-free People Tracking

Yang Zhao
GE Research
Niskayuna, New York, USA

Ming-Ching Chang
University at Albany, SUNY
Albany, New York, USA

Peter Tu
GE Research
Niskayuna, New York, USA

ABSTRACT

Recent radio frequency (RF) sensing techniques use a network of RF sensors to locate people without any carrying devices even at non line-of-sight environments. Model-based device-free RF sensing systems use statistical models to quantify the human presence and motion based on the received RF signal measurements. However, such methods often require fine tuning many model-dependent parameters to achieve sub meter accuracy. In this work, we propose to use deep neural networks together with visual tracking systems to effectively generate training data and learn a general model. Our method can automatically produce human motion and occupancy images from RF sensor network measurements without the need of manual RF parameter tuning.

KEYWORDS

Deep Neural Networks, Detection, Tracking, RF Sensor Network.

1 INTRODUCTION

As billions of sensors and embedded devices are deployed and interconnected ubiquitously, tremendous amount of data are available for artificial intelligence (AI) algorithms to find data patterns, predict trends, and make systems intelligent and autonomous. Deep Neural Network (DNN) based learning methods have been a roaring success in many fields solving real-world problems, including object detection/recognition, human activity recognition [3, 4] *etc.* In this work, we propose to apply DNN to an important Internet of Things (IoT) application — device-free people detection and tracking using radio frequency (RF) sensor network [5–7].

A network of low-cost, RF sensors can detect and track human motion and presence in real-time even at non line-of-sight (LOS) environments [5, 6]. Since this device-free technology does not require the use of visual sensors or any person-borne devices, it has great potential in many applications including security, smart facility, building energy management, emergency first response, search and rescue, *etc.* However, many state-of-the-art RF sensor network based people tracking systems require domain experts to configure and calibrate the systems before accurate detection and localization performance can be achieved. For example, the work of [7] requires an expert to manually tune six or more parameters for the kernel distance-based Radio Tomographic Imaging (RTI) system to achieve sub meter accuracy at a multipath-rich environment. To remove the human expert in the loop, we propose to use: (1) supervised deep learning methods together with (2) automatic labeled data generation using visual detection and tracking systems [2], to effectively learn a generic RF sensing model for device-free people detection and tracking.

2 METHODS

Figure 1 shows the overall architecture of the proposed intelligent people tracking system. Computer vision (CV) system can detect and track people in well-lit conditions at line-of-sight (LOS) environment. Recent visual object detection algorithms (YOLOv3, SSD, Faster-RCNN, or Mask R-CNN) can accurately detect and localize people when they are visible. On the other hand, RF sensing can augment visual sensing and compensate the invisible situations. We propose to combine the CV systems [2] and RF sensor network to build the next-generation indoor site-wide people detection and tracking system for occupancy estimation in a real-world setting. The combination of the two sensing modalities can increase coverage, bootstrap each other on improving detection accuracy and reducing errors. Furthermore, CV systems can be used to create ground truth data for automatic RF sensing parameter tuning.

For none-LOS (NLOS) environments, we can adopt existing NLOS experiments [5, 7] and design new experiments with CV systems, to obtain location ground truth data for data-driven parameter learning. With the learning of both LOS and NLOS environments, a generic model can be built to work in diverse settings and environments. We next describe the device-free RF tracking system, computer vision tracking system, and automatic RF parameter learning methods in greater details.

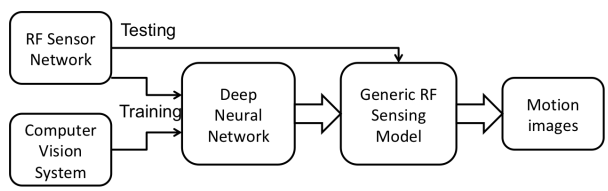

Figure 1: System Architecture.

2.1 Device-free RF tracking system

Unlike RFID localization system, device-free RF sensing system directly uses attenuation, reflection and scattering effect of human body on radio signal to detect and localize people, thus it works without the need of person-borne devices. In this paper, we focus on Kernel distance-based Radio Tomographic Imaging (KRTI) [7], a model-based device-free RF tracking system. The KRTI system uses low-cost radio transceivers as sensors to reconstruct an occupancy image (examples shown in Figure 2). Specifically, the KRTI algorithm estimates the occupancy image $\hat{\mathbf{x}}$ from RF measurements $\mathbf{d}$ as:

$$\hat{\mathbf{x}} = (W^T C_{\mathbf{n}}^{-1} W + C_{\mathbf{x}}^{-1})^{-1} W^T C_{\mathbf{n}}^{-1} \mathbf{d}, \tag{1}$$

where $C_{\mathbf{x}}$ is the covariance matrix of the occupancy $\mathbf{x}$ (human presence), and $C_{\mathbf{n}}$ is the covariance matrix of the measurement noise. The occupancy covariance can be modeled as [7]:

$$\left[\frac{1}{\sigma_n^2} C_{\mathbf{x}}\right]_{i,j} = \frac{\sigma^2}{\delta} \exp\left(-\frac{\|\mathbf{r}_j - \mathbf{r}_i\|}{\delta}\right), \tag{2}$$

where $\sigma^2 = \sigma_x^2/\sigma_n^2$ is the ratio of variance of occupancy σ_x^2 to the variance of noise σ_n^2 that is used as a regularization parameter, δ is a

correlation distance parameter, $\mathbf{r}_i$ and $\mathbf{r}_j$ are the center coordinates of the i-th and j-th pixels. § 2.4 will describe the effects of the covariance parameters σ_x^2 and δ *w.r.t.* the occupancy images. Once the occupancy images are estimated, people locations and counting can be directed estimated from the resulting heat map images.

2.2 Computer vision tracking system

Visual detection from low-cost cameras can provide accurate people localization in the LOS environments. We propose to leverage visual person detection and tracking system in two ways: (1) Computer vision system can be used to gather ground-truth data for automatic RF parameter learning. (2) The combination of visual detection and RF detection systems can jointly increase coverage and improve the person tracking accuracy and robustness. State-of-the-art visual detection methods such as the Faster-RCNN with ResNet as feature network provide a good combination and trade-off on performance and speed (over 80% mAP at 5 FPS). Real-time people tracking can be performed following a standard tracking-by-detection paradigm, based on Hungarian assignment for target association and Kalman filter for robust tracking [1].

We propose to extend our existing through-wall experiments [5, 7] by adopting visual people tracking to effectively generate a rich training dataset for general device-free RF model learning. For example, we can design and perform experiments on NLOS RF sensor and LOS camera deployments, using ground truth labeling produced from the automatic CV tracking system. By collecting a sufficiently large labeled dataset, machine learning can be applied for data-driven RF parameter optimization.

2.3 Automatic RF parameter learning

We propose a data-driven approach to automatically learn RF parameters that can generate people occupancy images $\mathbf{x}$ from RF sensor network measurements $\mathbf{d}$. Input to this process includes the training dataset $D = (\mathbf{d}_i, \mathbf{x}_i)_{i=1}^N$ obtained from both the visual tracking results and the NLOS experimental datasets. We will investigate two deep learning methods: (1) Deep Reinforcement Learning (RL), and (2) automated transform by manifold approximation (AUTOMAP) [8].

Deep RL: We propose to learn a policy that can optimize RF model parameters from a standard initialization. The policy will be updated via Q-learning that iteratively refines the Q-values (state-action pairs) by maximizing rewards, which can be estimated by calculating a loss measurement reflecting the distance between the current RF tracking heatmaps and the ground-truth visual tracking heatmaps.

AUTOMAP: This framework [8] will be used to directly learn a reconstruction mapping between the RF sensor data and the occupancy images. The RTI methods [5, 7] essentially solve an inverse problem. AUTOMAP recasts the solution of an inverse problem (*i.e.*, image reconstruction) as a data-driven supervised learning problem. This framework can be used to reconstruct human occupancy images from RF sensor data with labeled training data.

2.4 Preliminary experiments

We have setup the RF sensor network and performed experiments on person tracking. Initial results show the effects of different RF model parameters on the quality of reconstructed images. Figure 2 shows the resulting heatmap of a person walking along a pre-marked path at a site of 8×12 feet that are equipped with sixteen RF sensors. We choose the covariance model parameters σ_x^2 and δ as described in Section 2.1 to illustrate the effects of model parameters that affects the quality of the resulting heatmaps. Two human motion images are shown in Figure 2 with different parameter values. In general, we found that the images from the KRTI algorithm are less sensitive *w.r.t* the change of only one single parameter. However, results can have significant fluctuations when multiple parameters are changing all together.

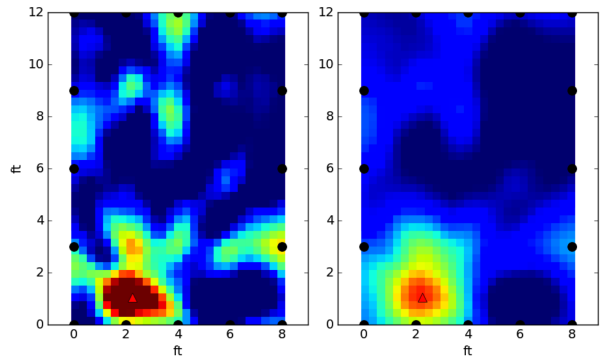

Figure 2: KRTI images of one person in a multipath-rich office environment (pixels with maximum values shown in red triangles are estimated person locations, left image from $\sigma_x^2 = 0.2$, $\delta = 5$; right image from $\sigma_x^2 = 0.1$, $\delta = 20$).

Observe in Figure 2 that large potential improvements can be done for the proposed RF parameter learning approach, particularly on the selection of model parameters. We expect this improvement can greatly advance the RF tracking usage in the complex multiple people scenarios with cluttered backgrounds.

3 CONCLUSION AND FUTURE WORK

We propose to combine RF sensing with visual tracking, where the framework can leverage deep neural networks to automatically generate a quality dataset for parameter learning. Early results of model parameters and performance analysis are shown on the KRTI system. We will continue to improve the proposed approach for device-free, multiple people RF tracking.

REFERENCES

[1] Ming-Ching Chang, Yi Wei, Nenghui Song, and Siwei Lyu. 2018. Video Analytics in Smart Transportation for the AIC'18 Challenge. In *CVPR Workshop on AI City Challenge*.

[2] J. Chen, M. Chang, T. Tian, T. Yu, and P. Tu. 2015. Bridging computer vision and social science: A multi-camera vision system for social interaction training analysis. In *ICIP*. 823–826.

[3] Lipeng Ke, Ming-Ching Chang, Honggang Qi, and Siwei Lyu. 2018. Multi-Scale Structure-Aware Network for Human Pose Estimation. In *ECCV*.

[4] Wenbo Li, Longyin Wen, Ming-Ching Chang, Sernam Lim, and Siwei Lyu. 2017. Adaptive RNN Tree for Large-Scale Human Action Recognition. In *ICCV*.

[5] Joey Wilson and Neal Patwari. 2010. Radio Tomographic Imaging With Wireless Networks. *IEEE Transactions on Mobile Computing* 9, 5 (May 2010), 621–632.

[6] Kristen Woyach, Daniele Puccinelli, and Martin Haenggi. 2006. Sensorless Sensing in Wireless Networks: Implementation and Measurements. In *WiNMee*.

[7] Yang Zhao, Neal Patwari, Jeff M. Phillips, and Suresh Venkatasubramanian. 2013. Radio Tomographic Imaging and Tracking of Stationary and Moving People via Kernel Distance. In *IPSN*. 229–240.

[8] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. 2018. Image reconstruction by domain-transform manifold learning. *Nature* 555, 7697 (2018), 487.